Tecnológico de Monterrey

Redes Bayesianas

Daniel Otero Fadul

Departamento de Ciencias
Escuela de Ingeniería y Ciencias

Redes Bayesianas

Una **red Bayesiana** es una estructura que nos permite representar dependencias entre distintas variables aleatorias. Este tipo de redes pueden representar, prácticamente, cualquier distribución conjunta, y en muchas ocasiones de manera muy concisa. Esta representación se realiza por medio de un **grafo acíclico dirigido**.

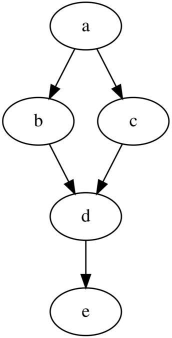

Figura: Un **grafo acíclico dirigido** es un tipo de grafo en el cual sus arcos tienen una dirección y que no posee ciclos que establezcan un camino cerrado entre un nodo, o vértice, y el mismo. Imagen tomada de tomada de `https://en.wikipedia.org/wiki/Directed_acyclic_graph`).

Redes Bayesianas

En el contexto de las redes bayesianas, estas se pueden considerar grafos acíclicos dirigidos que cumplen con las siguientes características:

- Cada nodo es una variable aleatoria, la cual puede ser discreta o continua.
- Un conjunto de arcos dirigidos conectan pares de nodos. Si hay un arco que va del nodo x al nodo y, se dice que x es el *padre* de y.
- Cada nodo x_i que tenga *padres* tiene asociada una distribución de probabilidad condicional $P(x_i|\text{padres}(x_i))$.

Redes Bayesianas

La topología de la red, el conjunto de nodos y arcos, especifica la forma en que se establecen las relaciones de causa y efecto entre las variables: las causas son los padres de los efectos. Normalmente, estas relaciones las establece un experto.

Una vez establecida la topología de la red Bayesiana, ya es posible determinar la forma de las distribuciones de probabilidad condicionales de cada variable dados sus padres. La combinación de las distribuciones de probabilidad de cada nodo es suficiente para obtener la forma de la distribución de probabilidad conjunta de todas las variables.

Redes Bayesianas

Como sabemos, gracias a la regla de la multiplicación, cualquier distribución de probabilidad conjunta se puede escribir de la siguiente forma:

$$\begin{aligned} P(x_1, x_2, \ldots, x_n) &= P(x_1)P(x_2|x_1)\cdots P(x_n|x_{n-1}, x_{n-2}, \ldots, x_1) \\ &= P(x_1)\prod_{i=2}^{n} P(x_i|x_{i-1}, \ldots, x_1). \end{aligned}$$

En el contexto de las redes Bayesiana, esta misma probabilidad conjunta será igual a

$$P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i|\text{padres}(x_i)).$$

Esto es válido siempre y cuando padres$(x_i) \subseteq \{x_{i-1}, x_{i-2}, \ldots, x_1\}$. Esta condición se puede satisfacer si la enumeración de los nodos es consistente con el orden implícito de la estructura del grafo. Por cierto, si la variable aleatoria x_i no tiene padres, tenemos que $P(x_i|\text{padres}(x_i)) = P(x_i)$.

Redes Bayesianas

Expresar la distribución de probabilidad conjunta como la multiplicación de las probabilidades condicionales ya mencionadas es posible si se cumple que

$$P(x_i|x_{i-1},\ldots,x_1) = P(x_i|\text{padres}(x_i)). \quad (1)$$

Esta condición permite que la red Bayesiana sea una representación correcta de la distribución de probabilidad conjunta si cada nodo es condicionalmente independiente de sus predecesores dado sus padres. La siguiente metodología garantiza esta condición:

1. **Nodos:** Determinar todas las variables del modelo. Una vez hecho esto, enumerarlas de tal forma que las causas precedan a los efectos.
2. **Arcos:** Para cada x_i:
   - Del subconjunto $\{x_{i-1}, x_{i-2}, \ldots, x_1\}$, escojer un conjunto mínimo de padres de tal forma que la Ecuación (1) se satisface.
   - Conectar cada padre por medio de un enlace con x_i.
   - Determinar la **tabla de distribución condicional** $P(x_i|\text{padres}(x_i))$.

Redes Bayesianas

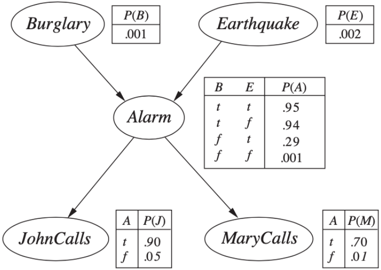

Figura: Un ejemplo de una red Bayesiana en la que se puede ver su topología y sus tablas de distribución condicional. E las tablas, las letras B, E, A, J y M hacen referencia a "Burglary", "Earthquake", "Alarm", JohnCalls" y "MaryCalls", respectivamente. Nótese que $P(\text{MaryCalls}|\text{Burglary, Alarm}) = P(\text{MaryCalls}|\text{Alarm})$. Por otro lado, las probabilidades de cada fila de las tablas deben sumar uno, ya que todas las variables son binarias, se omiten los valores que tomarían las distribuciones condicionales para los complementos de los eventos B, E, A, J y M. Imagen tomada de [1].

Una vez determinada la red podemos hacer cosas como la siguiente:

$$\begin{aligned} P(\bar{b}, \bar{e}, a, j, m) &= P(\bar{b})P(\bar{e})P(a|\bar{b}, \bar{e})P(j|a)P(m|a) \\ &= (0.999)(0.998)(0.001)(0.9)(0.7) = 0.000628. \end{aligned}$$

Redes Bayesianas

Figura: Una escena en la que Amy Santiago habla del problema de "Monty Hall" en la serie "Brooklyn 99". Imagen tomada de
`https://thelachatupdate.com/2018/11/06/the-monty-hall-problem-simplified/`.

Redes Bayesianas

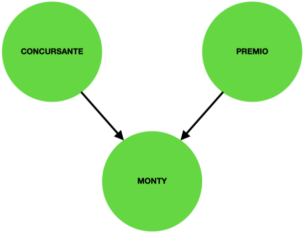

Figura: La red Bayesiana del problema de “Monty Hall”.

Dado lo que hemos visto, las variables aleatorias de este problema son “Concursante”, “Premio” y “Monty”, las cuales denotamos como C, P y M, respectivamente. Sean 1, 2 y 3 las puertas del problema de “Monty Hall”; estos números son los valores que toman las variables aleatorias. Nótese que las variables aleatorias C y P son independientes y que la variable M depende de C y P.

Entonces, tenemos que

$$
\begin{aligned}
P(C,P,M) &= P(C)P(P|C)P(M|C,P) \\
&= P(C)P(P)P(M|C,P).
\end{aligned}
$$

Además, las tablas de $P(C)$ y $P(P)$ son las siguientes:

	$C=1$	$C=2$	$C=3$
$P(C)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

	$P=1$	$P=2$	$P=3$
$P(P)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

En cuanto a la tabla de $P(M|C,P)$, tenemos que

	$M=1$	$M=2$	$M=3$
$P(M\|C=1,P=1)$	0	$\frac{1}{2}$	$\frac{1}{2}$
$P(M\|C=1,P=2)$	0	0	1
$P(M\|C=1,P=3)$	0	1	0
$P(M\|C=2,P=1)$	0	0	1
$P(M\|C=2,P=2)$	$\frac{1}{2}$	0	$\frac{1}{2}$
$P(M\|C=2,P=3)$	1	0	0
$P(M\|C=3,P=1)$	0	1	0
$P(M\|C=3,P=2)$	1	0	0
$P(M\|C=3,P=3)$	$\frac{1}{2}$	$\frac{1}{2}$	0

Para resolver el problema de "Monty Hall" es conveniente obtener la distribución de $P(P|C=1, M=2)$:

$$\begin{aligned} P(P|C=1, M=2) &= \frac{P(C=1)P(P)P(M=2|C=1,P)}{P(C=1,M=2)} \\ &= \frac{P(C=1)P(P)P(M=2|C=1,P)}{P(C=1)P(M=2|C=1)} \\ &= \frac{\left(\frac{1}{3}\right)\left(\frac{1}{3}\right)P(M=2|C=1,P)}{\left(\frac{1}{3}\right)\left(\frac{1}{2}\right)} \\ &= \frac{2}{3}P(M=2|C=1,P). \end{aligned}$$

Por lo tanto,

	$P=1$	$P=2$	$P=3$
$P(P\|C=1, M=2)$	$\frac{1}{3}$	0	$\frac{2}{3}$

Volviendo al ejemplo de Mary y John, supongamos que ahora queremos calcular $P(B|J=t, M=t)$. Si este es el caso, tendríamos que hacer lo siguiente

$$
\begin{aligned}
P(B|J=t, M=t) &= \alpha P(B, J=t, M=t) \\
&= \alpha \sum_{E} \sum_{A} P(B, J=t, M=t, E, A) \\
&= \alpha \sum_{E} \sum_{A} P(B) P(E) P(A|B, E) P(J=t|A) P(M=t|A),
\end{aligned}
$$

donde α es una constante. Se puede ver que, a pesar de que la red Bayesiana no es muy compleja, se requiere de muchos cálculos para poder obtener la distribución de probabilidad condicional $P(B|J=t, M=t)$. De hecho, para n variables Booleanas, la complejidad de este procedimiento es $O(n2^n)$.

Lo anterior es un ejemplo de **inferencia exacta** de una red Bayesiana. Es, obviamente, precisa, pero computacionalmente muy costosa. En general, los algoritmos que se utilizan para hacer estos cálculos se pueden dividir en dos categorías: los que realizan una inferencia exacta y los que obtienen una **inferencia aproximada**. Para mayor información, por favor, consultar [1].

BIBLIOGRAFÍA

1 Stuart Rusell,Peter Norvig, S. J., *"Artificial Intelligence: A Modern Approach"*, Cuarta Edición, Prentice Hall, 2020.

2 Charniak, E. *"Bayesian Networks without Tears"*, AI Magazine, 12(4), 50, 1991.